PAPER FOR PRESENTATION

SYMPOSIUM ON BIOAVAILABILITY
sponsored by the U.S. ENVIRONMENTAL PROTECTION AGENCY
and the SOCIETY FOR ENVIRONMENTAL GEOCHEMISTRY AND HEALTH

SEPTEMBER, 1990

November 1990

INTER-SITE COMPARISONS OF ENVIRONMENTAL LEAD UPTAKE

Allan H. Marcus and Alissa Bernholc

BATTELLE
Statistics and Data Analysis Systems
Research Triangle Park, NC 27709
(919) 549-8970

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 MATERIALS AND METHODS 5

3.0 COMPARISON OF TWO LEAD SMELTER COMMUNITIES -- A REGRESSION APPROACH 16

4.0 METHODOLOGICAL STUDIES OF EAST HELENA DATA SET USING STRUCTURAL MODELS 24

5.0 STRUCTURAL EQUATION MODELS IN THREE STUDIES WITHOUT HAND LEAD DATA 34

6.0 COMPARISON WITH THE MIDVALE STUDY 45

7.0 CONCLUSIONS AND DISCUSSION 53

REFERENCES 59

ABSTRACT

Comparison of lead exposures for children who live at different sites requires both comparability of data bases and comparability of statistical analysis methods. Data bases require comparable measures of environmental lead concentrations, common biological exposure indices such as venous blood leads, comparable protocols for sample collection, and child-specific mouthing behavior indices. Statistical models should have the same equation specifications, and should control confounding, multicollinearity, and surrogate variable measurement errors by use of coupled structural equation models. Site-specific adjustments for other covariates may be necessary. We illustrate solutions to these problems by comparing three epidemiology studies in lead smelting or mining communities: Kellogg-Silver Valley, Idaho; East Helena, Montana; Midvale, Utah. There may be some real site-specific differences in bioavailability of soil lead and in the strength of environmental pathways from soil lead to dust lead to blood lead. There were not significant differences in the relationship of household dust lead or hand lead to blood lead. Soil lead is a significant source of interior household dust lead in all communities, even when the direct pathway from soil lead to blood lead is not very significant.

INTER-SITE COMPARISONS OF ENVIRONMENTAL LEAD UPTAKE

Allan H. Marcus and Alissa S. Bernholc
Battelle Columbus Division
Research Triangle Park, NC

1.0 INTRODUCTION

There is a growing perception that not all forms of lead encountered by U.S. children in their environment are equally bioavailable (Chaney et al. 1988; Steele et al. 1990). This has some very important implications for site-specific health risk assessments, particularly with respect to setting clean-up levels for lead-contaminated soils. For example, lead in soil may come from several sources, including:

(1) deposition of airborne particles emitted by smelters or by other industrial sources;

(2) deposition of airborne fugitive dust emissions from mine wastes or from mill tailings piles;

(3) fill material from lead-contaminated sewage sludge, waste combustor residuals, etc.

(4) deposition of lead paint from deteriorating exterior lead-based paints on nearby buildings;

(5) deposition of residues from lead paint abatements;

(6) runoff from highways and other lead surfaces;

(7) fill material from other lead-contaminated sites containing any of the above sources;

(8) natural geological processes.

Thus the terms "soil lead" and "house dust lead" embrace a wide class of chemical and physical properties. Differences in the fraction of ingested lead absorbed into the blood are known, from animal experiments, to depend on particle size, chemical speciation, and dietary cofactors [see reviews by Chaney et al. (1988) and USEPA (1986, 1989)]. A number of studies on lead absorption by adult humans are qualitatively consistent with the animal studies, but do not provide definitive

quantitative extrapolation to young human children, the population at greatest risk. We are not aware of any clinical studies that can be used to quantify the soil lead uptake of young children. The only studies that are available appear to be observational studies of populations of lead-exposed children.

Some very important studies currently in progress are longitudinal studies in which changes in blood lead levels are observed following an intervention or abatement of a particular source of lead. These include the Superfund Soil Lead Abatement Demonstration Projects (SSLDAP) in Baltimore, Boston, and Cincinnati (Farrell et al. 1988; McIntyre et al. 1988; Clark et al. 1988). The interim report on baseline lead exposures has not yet appeared, and earliest results on the effectiveness of soil lead abatement in urban areas are not expected until June 1991.

Longitudinal studies on the effectiveness of paint lead abatement are now being carried out in New York City (Rosen et al. 1990). These studies include both blood lead and bone lead data as indices of childhood exposure, but do not contain sufficiently detailed information on lead pathways (soil, dust lead) or characterization of paint (e.g. lead chromate vs. lead carbonate) to assess bioavailability.

The least satisfactory source of data that is still useful is cross-sectional observational data. Typically, a random sample or a largely complete sample of children in some community is obtained. The information is as nearly child-specific as possible. Environmental data include observed lead concentrations in media to which the child is typically exposed, such as yard soil, dust at several locations in his/her dwelling unit, tap water lead, lead in outdoor air etc. Paint lead loadings at several interior and exterior locations on the dwelling unit are currently measured by X-ray fluorescence (XRF), but in earlier studies may be characterized as concentrations in paint chips. In some recent studies, lead loadings on childrens' hands is available. Many studies now include child behavioral assessments, parental involvement, and questions about other lead

sources in the child's environment (such as secondary occupational exposure, home lead hobbies and work, removal of lead-based paint). Blood is the response variable, and the regression coefficient between blood lead and soil lead concentration (also known as the blood-lead vs. soil lead "slope") is a useful relative aggregate index of exposure (Marcus and Cohen 1988).

A more general and useful method for interpreting cross-sectional blood lead studies has been introduced by Bornschein et al. (1985, 1988), the structural equation model. This is a form of pathway analysis that separates the direct and indirect effects of each lead exposure source. For example, regarding "soil lead" as a source of blood lead, we recognize both a direct pathway to blood lead (soil lead --> blood lead) from ingestion of exterior soil, and an indirect pathway in which soil lead contributes part of the house dust lead burden, which is subsequently ingested (soil lead --> dust lead --> blood lead). The relative importance of the source and the pathway nay thus be separated.

In this paper we will discuss the adequacy of the data and of the statistical methods of analysis for comparing the amount of soil lead uptake that can be expected for children living at various sites. The most important issues for data adequacy are the comparability of the measurements that are made at various sites, surrogate variable ("measurement error") problems with lead exposure indicators, and recruitment of the sample population. The most important statistical modeling problems are model specification, use of covariates and confounders, and estimation methods for lead pathway models. These are discussed in Section 2. In Section 3 we present results of blood lead regression analyses for two lead smelter communities visited by CDC in 1983: Kellogg, Idaho and East Helena, Montana. In Section 4, the role of model specification and statistical methodology are explored in detail for the East Helena data set. In Section 5, structural equations models are

compared for the Kellogg and East Helena sets, and in Section 6 for the Midvale, Utah site. The implications for cross-site comparison are reviewed in Section 7.

2.0 MATERIALS AND METHODS

2.1 DATA SET ADEQUACY

2.1.1 Data Sets Used In EPA Analyses and Subsequent Studies

Our discussion is based on the cross-sectional studies referred to in the USEPA Air Quality Criteria for Lead (1986), (Marcus and Cohen 1986), and the USEPA Exposure Analysis Methodology and Validation (1989). The soil or dust lead studies on which (USEPA, 1986, 1989) placed the greatest reliance were carried out in the following North American locations:

Active primary lead smelter:

- Bartlesville, Oklahoma, 1979.
- East Helena, Montana, 1983.
- El Paso, Texas, 1973.
- Kellogg, Idaho, 1974, 1983.
- Palmerton, Pennsylvania, 1979.
- Trail, British Columbia, Canada, 1975.

Primary zinc smelters:

- Ajo, Arizona, 1979.
- Anaconda, Montana, 1979.

Urban area with secondary lead smelter:

- Omaha, Nebraska, 1970's.

Urban area with other lead sources:

- Baltimore, Maryland, 1977-1990.
- Boston, Massachusetts, 1980-1990.
- Cincinnati, Ohio, 1980-1990.
- New Haven, Connecticut, 1974-1977.
- Charleston, South Carolina, 1975.
- Minneapolis-St. Paul, Minnesota, 1987.

There were also useful European soil or dust lead studies:

Urban areas:

- Rotterdam, The Netherlands
- Birmingham, England, U.K.
- Edinburgh, Scotland, U.K.

Secondary Lead Smelters:

Arnhem, The Netherlands, 1979.

Belgium, 1974-1978.

Old lead mining areas:

Derbyshire, England, U.K.

There have been a number of recent studies in former lead mining and smelting communities in the western U.S., including:

Leadville, Colorado, 1988.

Midvale, Utah, 1989.

Park City Utah, 1988.

Telluride, Colorado, 1987.

2.1.2 Comparability of Measurements

Problems in synthesizing the diverse data from these studies have been reviewed by USEPA and by Steele et al. (1990). The major problem is that soil lead has been sampled in cores at different depths and at different locations with respect to the child subjects, so that the predictiveness of a relationship found in any individual study may not be closely related to that in some other study where soil lead concentration was determined at some other depth or location. Even where sampling methods have been nearly identical, there may be significant differences in child lead exposure due to behavioral and cultural factors. For example, in warm climates there will be more opportunities for soil lead to enter house dust through open doors and windows, and more opportunity for children to play outside, than in cold climates. On the other hand, frequent use of air conditioners may reduce this pathway in warm climates. Frequency of hand cleaning or house cleaning may differ from one community to another. Awareness of hazards from lead-contaminated soil may also increase parental intervention in preventing childhood lead exposure, and this is likely to be greater in communities with large lead sources. The season when sampling is carried out is also important, since child blood lead levels are typically much

higher in summer than in winter, so that the blood lead vs. soil lead slope may indeed not be comparable if blood leads are measured at different seasons.

Soil lead and dust lead sampling methods have become increasingly standardized in recent years. The methods developed by investigators in the Baltimore, Boston, and Cincinnati soil lead abatement demonstration projects in progress reflect current "state of the art" approaches. Soil samples in cores at depths of 2 cm and 15 cm are taken at a uniform sampling grid close to the house, and other samples at the yard edge and in the child's play area. Dust samples are taken on floors in the child's indoor play areas and bedroom, and other locations. The sampling sites, indoors or outdoors, depend on the configuration of the dwelling unit and so cannot be completely uniform. Thus the composite "soil lead concentration" from one city may represent something different in different cities due to unintended differences in housing types. Even so, this is likely to be the most nearly comparable set of soil lead vs. blood lead data, by design, when these studies are completed.

Earlier studies that collected data generally comparable to the SSLDAP studies include preceding studies in Boston and Cincinnati, and studies carried out by the Cincinnati investigators in Telluride and Midvale. The studies carried out by the Centers for Disease Control in East Helena and in Kellogg in 1983 should be comparable to each other, and to similar studies in Park City. The Leadville study used very similar methods to the SSLDAP and the data should be comparable. All of the studies identified in this paragraph showed considerable sensitivity to capturing all of the major sources of childhood lead exposure. Differences in soil lead measurements and in dust lead measurement methods (e.g. dust loadings from wet wipes vs. lead concentration in vacuum cleaner samples) may need to be resolved.

2.1.3 Surrogate Variable ("Measurement Error") Problems

Even though sampling protocols for soil, dust, and blood lead are increasingly consistent, there is still some concern about the adequacy of these measurements as indicators of environmental lead exposure. In order to mechanistically model childhood soil lead exposure, one would have to have a detailed child behavior profile that estimates the length of time a child spends in each micro-environment (by season and weather), the amount of hand-to-mouth activity, and the frequency of hand cleaning and parental intervention. Then lead intake could be calculated accurately as the product of the soil lead concentration in each micro-environment, the amount of soil (or soil-derived house dust) consumed there, and interference with absorption of ingested soil lead from meals or other factors. This is not feasible.

Sometimes a reasonable surrogate variable can be constructed from interview data. For example, suppose the interviewer asks the child's caretaker "about how often does he/she put dirt or soil into his/her mouth? Once a week __ Once a month __ or practically never?" The responses could be coded, e.g. by monthly frequency (4, 1, 0), or a non-linear coding such as a geometric scheme (e.g. k*k, k, 1 respectively). Then the product of concentration by mouthing frequency may be more predictive of lead intake and blood lead than concentration alone.

The uncertainties introduced by use of surrogate exposure variables can have serious statistical consequences. These uncertainties are propagated (not resolved) when our understanding of environmental lead exposure is formalized in computer simulation models.

2.1.4 Recruitment of the Sample

The comparison of blood lead vs. soil lead slopes should also take into account the nature of the populations sampled. Differences in cultural or social aspects of child

behavior can mitigate blood lead vs. soil lead slopes. For example, different groups in the same community may have different patterns of lead exposure. No set of environmental variables alone can characterize those differences. The use of appropriate covariates may allow us to adjust lead exposure models, but does not guarantee an explanation of site-specific differences in lead uptake.

Other differences may be present among studies. For example, in a prospective longitudinal study, it may be appropriate to recruit only one sibling under age 7 from each family. The argument is that children from the same family may be more similar to each other than children from different families, thus show less variability in response than children from different families. On the other hand, it is often impossible to avoid enrolling all children whose caretakers would like them to be tested. A recent study in Leadville found no significant reduction in variability by including multiple siblings from the same family (Colorado 1989). The question of the population to whom a blood lead vs. soil/dust lead relationship can be generalized is always present.

2.2 ADEQUACY OF STATISTICAL METHODS

2.2.1 Model Specification

In modelling the relationship between blood lead and environmental lead, the primary concern is that the model should be consistent with our understanding of the biological and physical properties of lead metabolism. The EPA documents have stated that the relationship between blood lead and lead exposure is apparently nonlinear at sufficiently high levels of exposure, but that the near-linear biokinetics of lead in humans implies a near-linear relationship between blood lead and environmental lead at currently important levels of exposure. That is, denoting lead concentrations in blood, air, dust, soil, and water respectively by PbB, PbA, PbD, PbS, PbW, then for some

coefficients b_2, b_1 etc., the underlying relationship is approximately of the form

$$PbB = b_2 + b_1*PbA + b_2*PbD + b_3*PbS + b_4*PbW + etc.$$

(Equation 2-1)

The distribution of the response variable PbB and of the predictor variables PbA, PbD, PbS, PbW is generally highly skewed, often approximately log-normal. The usual assumptions for fitting linear statistical models require that, for fixed values of the predictors, the response variable is normally distributed, with the same variance at each set of (fixed) predictor values. This is often contradicted in observational data sets, in which variability tends to increase with increasing mean blood lead. This suggests that a better specification would be the multiplicative error model,

$$PbB = (b_2 + b_1*PbA + b_2*PbD + b_3*PbS + b_4*PbW + etc.)* exp(random error)$$

(Equation 2-2)

After taking logarithms of both sides of Equation 2-2, we have

$$\log(PbB) = \log(b_2 + b_1*PbA + b_2*PbD + b_3*PbS + b_4*PbW + etc.) + random error$$

(Equation 2-3)

This is a much more appropriate form for estimation of the regression coefficients for predicting geometric mean blood lead (GMB), GMB = exp(predicted log(PbB)). The sum of log-normally distributed components on the right-hand side of Equation 2-2 will also have a highly skewed distribution (not necessarily lognormal), but the logarithmic transformation of the sum on the right-hand side of Equation 2-3 will also tend to normalize the

distribution as well. Our regression analyses in Section 3 are based on Equation 2-3.

Because of the desire to use normally distributed variables in regression analyses and structural equation models, an alternative model is often used for curve-fitting,

$$\log(PbB) = \log(K_0) + K_1*\log(PbA) + K_2*\log(PbD) + K_3*\log(PbS) + K_4*\log(PbW) + \text{etc.}$$

(Equation 2-4)

Note that Equation 2-4 is very different from Equation 2-3, and that the regression coefficients in Equation 2-4 cannot be directly translated into those of Equation 2-3. Equation 2-4 is also physically implausible. Suppose that we set water lead levels to zero (PbW = 0) in a lead smelter community with elevated levels of PbA, PbD, and PbS. Then Equation 2-4 yields an estimated geometric mean blood lead of 0 μg/dl, which is absurd. Yet Equation 2-4 or its analogues have been used by many investigators without regard to the limited range over which it can be extrapolated. Equation 2-4 is similarly limited in its ability to estimate blood leads from soil lead cleanup levels, since exponentiating Equation 2-4 on both sides results in the prediction model for geometric mean blood lead,

$$GMB = K_0 * PbA^{K_1} * PbD^{K_2} * PbS^{K_3} * \exp(\text{etc.})$$

(Equation 2-5)

As a practical matter, the goodness of fit of log(PbB) is roughly similar for both Equation 2-3 and 2-4 in many observational studies, since the amount of variance in individual blood leads explained by either prediction model is rather low, about 30%.

Our recommendation is that the appropriate linear model specification, fitted by a nonlinear regression method, is much

preferred to the intrinsically nonlinear model 2-5 fitted by a linear regression method. The widespread availability of good non-linear regression programs in statistical packages for mainframes and personal computers should make this recommendation practicable for non-statisticians. As with any iterative computational procedure, plausible initial values for estimates are needed to start the search procedure for "optimal" parameter estimates.

2.2.2 Use of Covariables and Confounders

Comparison of blood lead slopes at different sites is facilitated if the regression or structural equation models contain exactly the same set of covariates. For example, children from low SES/low income families in lead smelter or mining towns may live in older parts of town closer to the lead source, and may have less parental control of lead exposure. If all of a group of highly correlated predictor variables are used in a regression model, the estimated coefficients may be statistically unstable. Furthermore, the role of intervening variables may be confused, since the direct and indirect effects of the blood lead predictors are not separable except using structural equation models.

2.2.3 Estimation Methods for Pathway Models

Model specification issues are particularly important in fitting pathway models for lead. The estimation and testing of causal pathway models that determine mean and covariance structures is known as structural equations modeling. Theoretical arguments for linear relationships among air lead, dust lead, and soil lead are presented by USEPA (1989). A typical set of relationships for a pathway model from air lead to blood lead is:

$$PbS = a_0 + a_1\ PbA + \text{other soil lead sources} \qquad \text{(Equation 2-6)}$$

$$PbD = c_0 + c_1\ PbA + c_2\ PbS + \text{other dust sources} \quad \text{(Equation 2-7)}$$

$$PbB = b_2 + b_1*PbA + b_2*PbD + b_3*PbS + b_4*PbW + etc \quad \text{(Equation 2-8)}$$

Additional terms could involve exterior or interior paint lead, secondary occupational exposure, and paint removal. For reasons given above, it would be convenient to fit this linear system after logarithmic transformation. There are relatively few programs that can be used to estimate parameters in an intrinsically nonlinear relationship, of which SAS/ETS PROC SYSNLIN (SAS 1988) is probably the most common.

We do not recommend fitting models that are linear in logarithms, even though this is commonly done. The logarithmic linear models have the same problems of interpretation described in Section 2.2.1. The linear models are not required to meet assumptions of normal distributions, unless one chooses to use maximum likelihood estimation methods that assume normally distributed variables. Since other methods are available for structural equation estimation, it is not necessary to deliberately misspecify the intrinsically linear model (2-6) to (2-8).

An alternative approach is to fit the linear system specification (2-6 to 2-8) without logarithmic transformation, but using a relatively robust estimation method that does not require normally distributed variables. The asymptotically distribution-free generalized least square (AGLS) implemented in the EQS package (Bentler 1989) often works well in practice. However, AGLS is essentially a "method of moments" procedure, thus requires large sample sizes to produce good results. We have found the EQS elliptical distribution generalized least squares method (EGLS) often provides good models, even in small samples.

Systems of equations should not be fitted independently, if possible, since response variables in some equations (2-6 and 2-7) are used as predictor variables in other equations (2-8). This violates a basic implicit assumption in linear and nonlinear regression programs, that each input or predictor variable or covariate be known exactly, "without error". "Measurement error" in the predictors almost always attenuates the size of the estimated regression coefficient (reduces it to a value that is on average closer to 0 than the true value); see (Fuller 1988) for a comprehensive review. Correlated "measurement errors" in confounded covariates (such as dust lead and soil lead) could even reverse the apparent direction of the blood lead vs. environmental lead relationship (Kupper 1984). Programs for equation system estimation that allow specification of "measurement errors" are recommended.

3.0 COMPARISON OF TWO LEAD SMELTER COMMUNITIES -- A REGRESSION APPROACH

3.1 INTRODUCTION

The multiple linear and nonlinear regression analyses relate the concentration of lead in whole blood (PbB, ug Pb per deciliter of blood) to the concentration of lead in many environmental media such as lead in air (PbA, $\mu g/m^3$), in household dust (PbD, $\mu g/g$), and in soil (PbS, $\mu g/g$). In the remainder of this section we will examine only the effects of environmental lead on children of age one to five years.

The 1983 CDC studies in Kellogg, Idaho (1983a) and in East Helena, Montana (1983b) are particularly interesting because they represent two aspects of the problems of lead contamination in the vicinity of a primary lead smelter. The Bunker Hill smelter in Kellogg was closed in 1981, so that the Idaho values represent a more typical situation for cleaning up contaminated soils. The East Helena smelter was in operation in 1983 when the CDC survey was carried out (and still is), and it is known that the primary source of lead, atmospheric deposition, had remained at relatively similar levels for at least 5 or 6 years preceding the study. Thus the East Helena situation represents a long-term relatively equilibrated exposure with children in the most heavily exposed zone (PbA of 3 to 4 $\mu g/m^3$) being significantly at risk for elevated blood lead.

A multifactorial analysis of the East Helena data has been developed (Johnson and Wijnberg 1988; Marcus and Holtzman 1988), using a linear total exposure methodology that is consistent with the USEPA (1986) results. The metabolism and biokinetics of lead absorption, distribution, and excretion are sufficiently linear at the low-to-moderate blood lead levels of interest here (most less than 25 $\mu g/dl$) that a linear total exposure model appears justifiable. The distribution of PbB is highly skewed, so that a logarithmic transformation of both PbB and total exposure is usually required, thus intrinsically

near curve-fitting methods. An alternative analysis of the
chilling and Bain 1988) emphasized soil lead and excluded
of household dust lead (PbD) and air lead (PbA). Their
ses also used a statistical model that was linear in the
ithms of the exposure variables, resulting in excessively
curvature and in biologically implausible extrapolations to
low concentrations.

MATERIALS AND METHODS

The data consisted of household-specific and child-
fic information on N = 200 children in Idaho and N = 396
lren in Montana. The measurements are described in detail in
lling and Bain 1988; CDC 1983ab). The variables used in our
ses are PbB for each child, PbD and PbS for each house, PbA
n three concentric rings centered on the smelter (Area 1:
ln 1.6 km; Area 2: 1.6 to 4 km; Area 3: more than 4 km from
smelter). Other variables included the presence of a smoker
ne household (SMOKER = 1 if present, 0 otherwise), and income
l (LOWINC = 1 if less than $10,000; 0 otherwise).

Missing values for PbD and PbS were dealt with by
additional variables to estimate the average blood lead
ement of the missing cases. That is, if PbD were missing,
was assigned the value 0 and a variable PbDMis set to 1. If
were not missing, then PbDMis was set to 0. Likewise, PbSMis
d missing PbS.

The initial full model tested was

$$\ln(PbB) = \ln(b_2 + b_1\ SMOKER + b_2\ LOWINC + b_3\ PbA + b_4\ PbDMis + b_5\ PbD + b_6\ PbSMis + b_7\ PbS) \quad \text{(Equation 3-1)}$$

full model in Equation 3-1 fitted the Montana data, but the
ficient for PbA was exceptionally unstable for the Idaho
, the PbSMis coefficient nonestimable as no soil leads were
ing. The Idaho model thus had fewer parameters than the

Montana model. Note that the parameters b_5 and b_7 are the PbD and PbS slopes from the nonmissing data, whereas b_4 is the average blood lead increment for children with PbD missing and b_6 the average PbB increment for children with PbS missing.

The other variables used in Schilling and Bain (1988) included age, which could be dummy-coded (e.g., for AGE = 1, AGE = 2 etc.) or fitted by a quadratic polynomial. The age-related terms should have properly contributed multiplicatively, as there would be little lead increment from age alone if no PbA, PbD, PbS were present. That is, one way of including age effects would be to use a term like

$$b_{51}\ \text{PbD*(AGE=1)} + b_{52}\ \text{PbD*(AGE=2)} + b_{53}\ \text{PbD*(AGE=3)} + b_{50}\ \text{PbD*(AGE > 3)} \qquad \text{(Equation 3-2)}$$

instead of b_5 PbD, and likewise for PbDMis, PbS, PbSMis, where age-dependent changes related to hand-mouth and play activity are important. Schilling and Bain (1988) also found that the child's play outdoor hours (CPOHR) was somewhat predictive. We see little reason for using this variable as a main effect. That is, it might be used by b_7 PbS*CPOHR.

The air lead variable PbA is somewhat confounded with AREA in these studies, which is the reason we have omitted it. As there may be other differences among the study areas, this could also be studied by using b_{31} (AREA = 1) + b_{32} (AREA = 2) instead of b_3 (PbA). Large differences in PbB response by AREA that are not accounted for by PbA, PbD, PbS, or LOWING might be difficult to explain.

Other variables, such as lead painted housing or secondary occupational exposure, could also be tested. The log transform of PbB is known to be normalizing and variance-stabilizing. The log transformation of the left side of Equation 3-1 (total exposure to lead is not required by the usual assumptions of nonlinear multiple regression, but is also likely to be normalizing. The distributions of PbA, PbD, PbS are

correlated and highly skewed, hence the total exposure term is also likely to have a highly skewed distribution that is nearly normalized by taking logs. This is more defensible than the use of a log-log linear model

$$\ln(PbB) = \ln(d_0) + d_1 \ln(PbA) + d_2 \ln(PbD) + d_3 \ln(PbS) + \text{etc.} \qquad \text{(Equation 3-3)}$$

The problem with Equation 3-3 is that the predicted PbB = 0 if any input PbA = 0, PbD = 0, or PbS = 0. Even in a less extreme case, the PbB vs. PbS slope is nearly infinite for low PbS, flat for large PbS. The back-transformed version of Equation 3-3 is a power function model for the geometric mean (G.M.) PbB,

$$\text{G.M. PbB} = d_0 \, PbA^{d_1} \, PbD^{d_2} \, PbS^{d_3} \exp(\text{etc.})$$

whereas the model for Equation 3-1 is

$$\text{G.M. PbB} = b_2 + b_1 \, \text{SMOKER} + b_2 \, \text{LOWINC} + b_3 \, \text{PbA} + \ldots \qquad \text{(Equation 3-4)}$$

There is no plausible biological justification for a completely multiplicative model such as in Equation 3-3, whereas Equation 3-4 is consistent with linear pharmacokinetics for lead (USEPA 1986) at relatively low exposures. Either model should be able to give a good description of the data over the observed range, but the linear model may be more appropriate to estimating the effects of drastic lead reductions in air or soil (PbA close to 0, PbS close to 0).

3.3 RESULTS

The results of the first set of analyses is shown in Table 3-1. The models both have large residual standard deviations, 0.467 in log units for Idaho and 0.445 for Montana,

TABLE 3-1. COMPARISON OF LOG-LINEAR REGRESSION MODELS FOR CHILD BLOOD LEAD

Variables	CDC/ATSDR 1983	
	Kellogg	E. Helena
AIR LEAD	----	0.65*
		(0.23)
DUST LEAD	1.36*	1.26*
	(0.47)	(0.35)
SOIL LEAD	0.86*	1.54*
	(0.22)	(0.64)
MISSING DUST	1.42	1.85*
	(1.27)	(0.48)
MISSING SOIL	----	-1.24*
		(0.58)
LOW INCOME	1.88	-0.18
	(1.01)	(0.69)
INTERCEPT	7.69	5.37
	(0.96)	(0.38)

NOTE: Estimated standard errors in parentheses. Statistically significant variables marked by asterisks.

but are at least as small as those in Schilling and Bain (1988). The only nonsignificant terms are b_2 (LOWINC) in both states, and b_4 (PbDMis) in Idaho. The PbD coefficient b_5 is almost the same in both locations, 1.361 ± 0.470 (s.e.) μg/dl per 1000 ppm in Idaho and 1.263 ± 0.353 μg/dl per 1000 ppm in Montana. The children who had missing PbD values had an average dust lead increment of 1.419 ± 1.267 μg/dl in Idaho and 1.854 ± 0.483 μg/dl in Montana. The Montana children with missing PbS had a PbB increment of b_6 = -1.241 ± 0.582 μg/dl. The PbS slope b_7 showed an interesting difference, 0.860 ± 0.225 μg/dl per 1000 ppm in PbS in Idaho and 1.537 ± 0.638 μg/dl per 1000 ppm in Montana. This difference is not statistically significant, but it is suggestive. The "background" level b_2 is significantly lower in the Montana children, 5.306 ± 0.374 μg/dl vs. 7.681 ± 0.970 μg/dl in Idaho. The SMOKER effect b_1 is also higher in Idaho, 2.796 ± 0.935 μg/dl vs. 1.351 ± 0.392 μg/dl in Montana, but this term together with b_2 for LOWINC may be a surrogate socioeconomic factor.

Both Idaho and Montana studies have substantial predictive power for population geometric means, but very modest predictive power for any given individual. The usual assessment criterion is the coefficient of determination R^2 i.e., the fraction of variance about the mean (SSM) that is attributable to regression,

$$R^2 = 1 - SSE/SSM$$

where SSE is the residual sum of squares of the model. For Idaho, R^2 = 0.283; and for Montana, R^2 = 0.256.

3.4 DISCUSSION

The multiple regression model fitted here was a transformed linear total exposure model. While the model does not significantly improve the dose-response description, it offers much greater biologically plausibility, interpretability,

and low-exposure extrapolatability than does a log-log model. At low to moderate exposure levels in each lead source, the interindividual differences in lead absorption and the individual differences in nonattributed exposure sources ("background") are relatively much more important. Thus the residual GSD for these children

$$\text{residual GSD} = \exp[\text{std.dev.residual } \ln(\text{PbB})]$$

is very large, GSD = 1.596 for Idaho and GSD = 1.561 for Montana. Thus the use of the models for estimating individual PbB is not recommended. The current PbB vs. PbD slopes are on the low end of the range in USEPA Criteria Document, 1.361 for Idaho and 1.263 for Montana, and the current PbB vs. PbS slopes are similarly low, 0.860 for Idaho and 1.537 for Montana.

The soil lead relationship in the 1983 study for the Idaho Silver Valley is similar to that of the 1974-1975 study, reestimated in USEPA (1986). The PbB vs. PbS slope there was 1.10 ± 0.14. Household cleanliness was also a surrogate factor for PbD, with a strong effect during the era when atmospheric deposition was very high. Because the preponderant source of lead exposure was the then-noncompliant smelter, there was much less relative variability in the blood lead vs. environmental lead relationship, with a GSD of about 1.30.

Neither lead smelter community is unaware of the health hazards of undue lead exposure. Public awareness and public health education efforts must have increased parental vigilance, on the whole, so that opportunities for contact with leaded soil and dust during play were reduced, and the frequency of hand washing, house cleaning and other hygiene practices were increased. All of these serve to attenuate the apparent relationship of blood lead to environmental lead. Public health professionals are to be commended for the apparent success of these efforts. But it may be still more effective to remove the

sources of lead contamination, thereby protecting the unaware children as well as the vigilant ones.

The PbB vs. PbS slope is lower (not quite at 5% significance) in Idaho than Montana. This could be due in part to greater awareness in Idaho, or to reduced biological availability of the particles in the Idaho surface soils. The Idaho leaded soils may have different physical or chemical properties, possibly having resided longer without replenishment from the source smelter.

4.0 METHODOLOGICAL STUDIES OF EAST HELENA DATA SET USING STRUCTURAL MODELS

4.1 METHODS AND MATERIALS

In the East Helena study, outdoor environmental lead was measured in the air (PbA, in $\mu g/m^3$) and soil (PbS, in mg/g) surrounding the homes of 390 children, aged one to five years. Indoor lead sources were estimated by measuring lead in house dust (PbD, in mg/g), and by a variable indicating the presence of lead paint in the houses (Paint). The biological measure of lead exposure was lead concentration in blood (PbB, in $\mu g/dl$). Age was also included as an independent variable, since blood lead levels have been shown to be age dependent. Because the age dependence is not a linear one, we used indicator variables for age, where AGE1=1 if age = 1, AGE2=1 if age = 2, AGE3=1 if age = 3, and AGEj = 0, j=1,2,3, otherwise.

Structural equations models used in the Cincinnati studies were different than those used here. In the Cincinnati studies, the air lead concentrations were moderate to low, and therefore air lead was not included in the models. However, in East Helena the air lead concentration was much higher (averaging 3-4 $\mu g/m^3$ at homes near the smelter), and direct exposure to air lead was found to make a significant contribution to blood lead levels, in addition to its indirect contributions to soil and dust lead.

Since the distribution of many environmental measures, including lead, are highly skewed, methods must be used to handle analyses of non-normal data. A common method is to log-transform the data to bring them closer to a normal distribution and then perform analyses on the transformed data. Log-log multiple regression models for blood lead were fitted by CDC (1986), Schilling and Bain (1988). There are two drawbacks to that method: 1) sometimes the transformation does not create normal variables, and 2) the relationship may actually be a linear one

of untransformed variables and not of transformed variables (Marcus and Cohen 1989).

A different alternative is to use statistical methods which either do not assume normality or which are robust against violations of normality. These multivariate methods have been only recently developed. Also, like other nonparametric methods, they have less power than parametric methods and therefore require a larger number of observations when the underlying distribution is in fact normal. However, parametric (e.g., maximum likelihood and elliptical) methods may fail badly if the distribution assumptions do not hold.

We have examined the data in several ways, using the EQS structural equations program developed by Bentler (1989). This program has the capability of performing generalized least squares, maximum likelihood, elliptical, and arbitrary distribution analyses. Since generalized least squares and maximum likelihood tests assume normality, a criterion not met in this study, we report and compare the elliptical and arbitrary distribution analyses for both nontransformed and log transformed data. Elliptical theory requires the distribution of variables to be symmetric, but not necessarily normal. It also assumes that all the variables have the same measure of kurtosis. Arbitrary distribution theory makes no assumptions about the shapes of the distributions. However it was found to have a significant but negligible bias in small samples (Browne, 1984). Data requirements for model fitting are shown in Table 4-1 for each method.

One drawback discovered in the EQS program is that it converts all missing values in raw data to zeroes. There were only two variables with missing values, PbS and PbD, but they are essential to any model. In order to circumvent this problem, we first performed an analysis using only complete observations. About 60% of the sample, or 229 children, were thus included.

We also devised a model-free method to impute missing values. Each city block in East Helena was numbered, and the

TABLE 4-1. SOME STRUCTURAL EQUATIONS MODELING METHODS AVAILABLE IN EQS

Methods	Distribution Assumptions	Data Requirements
Generalized Least Squares (GLS)	Normally Distributed Variables	Raw Data or Matrix Covariance
Maximum Likelihood (ML)	Normally Distributed Variables	Raw Data or Matrix Covariance
Elliptical Generalized Least Squared (EGLS)	Variables symmetrically Distributed with Equal Kurtoses	Raw Data Only
Arbitrary or Distribution-Free Least Squares (AGLS)	No Distribution Assumptions	Raw Data Only

numbered block location of each child was included in the data. We assumed that households located near each other would have similar lifestyles, and thus similar soil and dust characteristics. We therefore assigned the calculated geometric mean PbS and PbD value of the block to any observation missing those values. This brought our sample size up to 366, or about 94% of the sample.

This method was similar to the way air lead values were imputed. Air lead values were assigned on a block-by-block basis based on a smooth interpolation of average yearly concentrations from monitoring sites. This was based on our assumption of homogeneous mixing of air within a limited area. Comparable smoothing of soil and especially dust lead is less plausible.

Because the data were not normally distributed, and a linear relationship was assumed, we first found the best plausible linear model with nontransformed variables using asymptotic distribution free estimates. We then used that model as a basis for comparison with log transformed data.

In our search for the best model, we removed all nonsignificant relationships in a stepwise procedure and tested the fit. Unfortunately, when all nonsignificant relationships were removed, the model was saturated, and no goodness-of-fit test could be performed, i.e., the fitted covariance matrix of the variables exactly matched the observed covariance matrix. Therefore, in order to compare the linear and log models, we needed to include one nonsignificant relationship. We chose the most significant of them. The inclusion of this relationship had little effect on the size of the coefficients found in the saturated model. We compared the saturated and unsaturated models for both linear and log-transformed data.

4.2 RESULTS

We will not shown the detailed results here. They are presented in Marcus and Bernholc (1989). The models needed for intersite comparisons are given in Sections 5 and 7. Figure 4-1 shows the unsaturated structural equation diagram fitted by the AGLS on nontransformed variables and used as a basis of comparison. Only the contribution of house-paint lead was not significant in the larger data set (with missing values imputed). We were surprised that the house-paint lead did not directly contribute significantly to either dust lead or blood lead. However, it did contribute to soil lead in the smaller data set. We assume that the exteriors of most houses whose interiors where painted with lead paint were also painted with lead paint, and that chipping and erosion occurs more heavily on the outside of houses. Another possibility is that the older houses (i.e., more lead paint and/or more paint chipping) were located closer to the smelter, where both the air and soil would have a higher lead content. This hypothesis is strengthened by the finding of a significant correlation between lead in air and lead in housepaint. However, the relationship between housepaint lead and soil lead cannot be explained totally by air lead.

Three equations were necessary to define this model:

$$PbB = a_0 + a_1\ PbD + a_2\ PbS + a_3\ PbA + E_1 \qquad \text{(Equation 4-1)}$$

$$PbD = b_2 + b_1\ PbS + b_2\ PbA + E_2 \qquad \text{(Equation 4-2)}$$

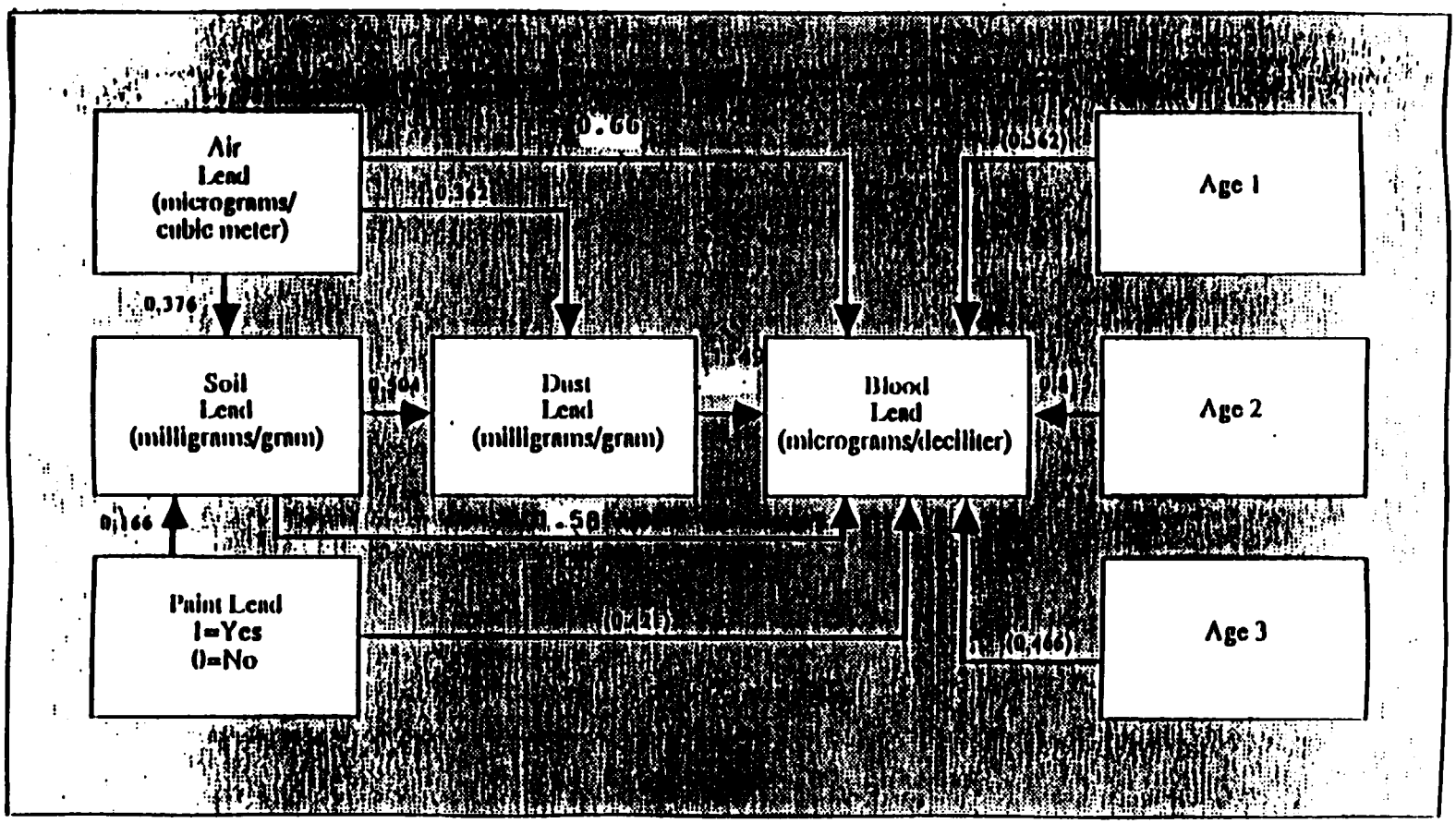
$$PbS = c_0 + c_1\ PbA\ (+ c_2\ Paint) + E_3 \qquad \text{(Equation 4-3)}$$

The coefficients and their standard errors are different for each method used, as well as for each data set and for raw vs. log transformed variables. A summary of model goodness of fit is given in Table 4-2.

TABLE 4-2. GOODNESS OF FIT STATISTICS FOR EAST HELENA ANALYSES

Data Set N	Data	Method	x^2	df	P
Complete 229	Raw	ERLS	1.801	2	0.406
		AGLS	3.574	2	0.167
	Log	ERLS	4.44	2	0.108
		AGLS	5.96	2	0.051
Expanded 366	Raw	ERLS	0.199	2	0.905
		AGLS	0.415	2	0.813
	Log	ERLS	3.983	2	0.137
		AGLS	5.306	2	0.070

FIGURE 4-1. ENVIRONMENTAL PATHWAYS FOR LEAD IN EAST HELENA CHILDREN
(Coefficient Estimated by Structural Equation Model)



In all three equations, the coefficients and standard errors are not affected greatly by the methodology used. Neither were they greatly influenced by including Paint in the model. The largest differences are found when comparing raw data to log-transformed data, and when comparing the smaller data set with only complete observations to the larger data set with imputed values for soil and dust lead.

In Equation 4-1, the effect of PbS on PbB, in the smaller data set, lost significance in the log transformed model, and the effect of PbA gained significance. Although this pattern is similar in the larger data set, it is not as drastic.

When comparing the smaller data set to the larger one, the effect of PbD on PbB seems to have lost some significance, although it remained an important contributor to PbB. The effect of both PbS and PbA on PbB gained significance. The reason for this dramatic shift of the coefficients from the smaller data set to the larger one is not apparent. Perhaps the method of imputation was not an accurate one. Or, the smaller data set may not be representative of the population (i.e., it might be biased in some way).

In Equation 4-2 all relationships are comparable and highly significant. The effect of PbS on PbD gained significance in log transformed data in both the saturated and unsaturated models for both data sets. However the effect of PbA on PbD lost significance for the log transformed data in the smaller data set, and gained significance in the larger data set. Both PbS and PbA were of higher significance in the larger data set.

Although PbA has a very significant effect on PbS, this significance is nearly doubled for log transformed data. Also, Paint seems to have a significant effect on PbS in the smaller data set, but not in the larger one. Possible reasons for this were stated in reference to Equation 4-1.

One can use goodness-of-fit statistics to test the fit of the model. The fit will depend upon the data set used, the type of data (raw or log-transformed), and the method of

estimating the coefficients. Note that the larger the p-value, the better the fit.

4.3 DISCUSSION

We have shown that different model specifications and statistical treatments of environmental lead data can lead to strikingly different conclusions about the relative importance of environmental lead pathways for children. A linear total lead uptake model is biologically plausible, at least at the lower end of the lead concentration ranges that are currently of interest. The linear specifications of the model that we use here are consistent with earlier USEPA studies using multiple linear and nonlinear regression methods (USEPA 1986, 1989). The asymptotically distribution-free methods in the EQS program allow estimation of model coefficients, even though the variables have highly skewed distributions. Although the coefficients obtained via ERLS were surprisingly similar to those obtained by AGLS, the significance of these coefficients differed. The ERLS method is much more restrictive, and since the data does not seem to meet these restrictions, the AGLS method is more valid.

The explicit inclusion of measurement errors E_1, E_2, and E_3 should greatly reduce the bias or attenuation of the estimated coefficients (Fuller 1987). Estimation of the lead coefficients as a coupled system should be more informative than separate estimation of the component equations.

The estimated standard deviations of E_1, E_2, and E_3 are about 4 μg/dl for blood lead, about 500 ppm for dust lead, and about 800 ppm for soil lead. The variability in E_2 and E_3 may indicate in addition to measurement error, the extent to which the measured variable, such as composite soil lead from front and back yards, may differ from the average lead concentration in the soils actually ingested by the child. Analyses with disaggregate measurements may suggest a very different weighting. For example, back yard soil concentrations might typically have twice the weight or influence than front-yard concentrations in

predicting blood lead, but back yard soil lead concentrations might have only half the influence of front yard soils in estimating living room dust concentration (which also affects blood lead). We have no way of knowing that such is the case, but this should be investigated in the forthcoming Superfund Soil Lead Demonstration Abatement Project or SSLDAP (Bornschein et al. 1988).

5.0 STRUCTURAL EQUATION MODELS IN THREE STUDIES WITHOUT HAND LEAD DATA

5.1. DATA SETS AND MODEL CONSTRUCTION

Most of the recent prospective lead abatement studies have used the child's hand lead burden as the most proximate index of childhood lead exposure from soil and dust lead. We here compare the two earlier CDC studies in Kellogg and East Helena, in which no hand leads were collected, with a more recent study carried out at Midvale, Utah. This will allow us to assess the role of statistical methodology and structural equations model formulation in a variety of situations in which the data bases contain relatively similar variables. We may then test the similarity of the models by formal statistical significance tests of the equality of corresponding regression coefficients.

The role of hand lead in these models is played by the product of household dust lead concentration multiplied by a normalized index of total mouthing behavior. The cross-product is therefore not a linear function of dust lead, although they are strongly correlated. In the Kellogg and East Helena studies, the mouthing behavior index was constructed from four binary indices and four three-point indices whose levels were "a lot", "once in a while", "almost never". We coded the binary indicators as 0 or 1, and the ternary indicators as 2, 1, 0 for each response. In the Midvale study, the ternary indicators were more directly related to frequency of consumption, with responses "practically never", "about once a month", and "about once a week" coded as 0, 1, and 4 respectively. Investigation of the most predictive weighting and coding of mouthing behavior indices would be useful. We summed the indices and divided the sum by its average in the whole data set, without regard to age or missing values of dust or soil lead. Thus the surrogate variable for hand lead burden is * (dust lead concentration) * (frequency

of all mouthing), where the mouthing frequency has been standardized as above.

The air lead concentrations were high only in the 1983 East Helena data set. Our reconstruction of air lead concentrations at each block in East Helena was possible only because of a rather dense network of air lead monitors at various sites in East Helena during 1981-1983. The CDC study in East Helena divided the houses into three concentric circular areas: Area 1 within 2.25 miles of the smelter, Area 2 from 2.25 to 5 miles, and Area 3 beyond 5 miles. The empirical air lead estimates showed a highly non-concentric pattern that closely matched the pattern of soil and dust lead concentrations. In order to facilitate comparison with the other two sites, we retain these three areas and assume that the dummy variables AREA1 (Area 1) and AREA2 (Area 2) are surrogate indices of exposure from air lead. The air lead levels around the inactive lead smelter in Kellogg were relatively low in 1983, and the three CDC areas correspond to the communities of Smelterville (AREA1), Kellogg (AREA2), and the more remote Pinehurst district as a baseline. The detailed location of the houses sampled in Midvale were also available, but there were at least two potential lead sources, an inactive smelter and some large tailings piles, so that we used a general quadratic model rather than a concentric pattern for spatial dispersion of lead in developing our Midvale model.

The use of AREA1 and AREA2, or X and Y, were not the only site-specific or idiosyncratic variables used in these models. Socio-demographic factors play an important role in blood lead levels. We used CDC's 3-level index of income for the Kellogg and East Helena models, but the more detailed Hollingshead scale for socio-economic status (SES) reported in Midvale. We likewise excluded information on lead paint, as this was missing or believed to be somewhat unreliable. The appropriate paint lead surrogates for the Midvale data were categories of house age (19th Century, post-War).

The models were otherwise completely comparable. The response variables in the four equations were blood lead, soil lead, dust lead, and the hand lead surrogate variable of dust lead times standardized frequency of mouthing behavior. The general scheme was:

(area, location, or air lead) --> soil lead

soil lead + air lead --> dust lead

dust lead + mouthing + (age) + income --> hand lead
(surrogate is dust lead X mouthing)

soil lead + dust lead + hand lead + air lead + (age) +
socio-economic status or income --> blood lead

These equations reflect plausible a priori hypotheses about causal processes and mechanisms. Similar models have been derived for other sites by the Cincinnati investigators. Therefore, the significance levels from the chi-squared tests for model adequacy may be taken at face value. All of the models described below were fitted initially by GLS estimates, which were improved by the EGLS method, and the final parameters estimated with the asymptotically distribution-free AGLS method. The AGLS models are all adequate at the usual nominal 5% significance level.

5.2 RESULTS

Our final effort to develop a plausible model for environmental lead pathways used the area-wide average air lead concentration. Even though air leads are currently very low in Kellogg, the low air lead concentration was a very useful and predictive surrogate for the previously much higher air lead levels, possibly because of fugitive emissions from smelter waste piles and from reentrainment of remaining pockets of fine surface soil and dust particles. Thus, the air lead regression coefficients for Kellogg are much higher than elsewhere, but the

predictive value of these concentrations should not be ignored. The East Helena air leads were high, and very predictive. Results are shown in Tables 5-1 through 5-5.

Table 5-1 shows that the AGLS model provides a satisfactory fit to the data, and that the blood lead equation is not the major source of lack of fit. Table 5-2 shows the blood lead models in more detail. Note that the hand lead surrogate coefficients are almost identical and highly significant, and that the difference between Kellogg and East Helena dust lead coefficients is not significant. Table 5-2 also shows that interior dust lead contributes strongly to blood lead. Both dust lead and dust lead X mouthing frequency are forced into the equation, and both are statistically significant in AGLS models. We could re-express this as

$$\text{blood lead} = b_1\,(\text{dust lead}) + b_2\,(\text{dust lead})\,(\text{mouthing}) + \ldots$$

However, the <u>total dust lead</u> coefficient varies from child to child: b_1 for a child with no perceived mouthing behavior, $b_1 + b_2$ for a child with average perceived mouthing behavior, $b_1 + 2\,b_2$ for a child with twice the average amount of mouthing etc. Income is not a significant predictor of blood lead, and age is significant only for the East Helena children. The blood lead vs. soil lead coefficient is significantly lower in Kellogg than in East Helena. Air lead levels, <u>characterized as area averages rather than household-specific values</u>, were not predictive in East Helena, but were highly predictive in Kellogg where air lead levels in 1983 were much lower. This underscores the importance of exposure indices that are specific to the individual subject, since the detailed model for East helena developed in the preceding section found that air lead on a block average basis provided a good predictor, over and above soil lead and dust lead levels. The air lead levels in the Idaho study were believed to be too low to provide much additional information, even if there

Table 5-1. Comparison of Structural Equation Models
GOODNESS OF FIT

	GLS		AGLS	
	Kellogg	E. Helena	Kellogg	E. Helena
Loss Function	0.19768	0.11475	0.05319	0.07875
Chi-Squared	27.873	26.508	7.500	18.190
DF	12	12	12	12
P	0.00577	0.00909	0.82290	0.11004
Largest Residuals in Variance or Covariance	VAR(PBS)	VAR(PBD* MOUTHS_ALL	VAR(PBD)	VAR(PBD* MOUTHS_ALL)
	COV(PBS,PBD* MOUTHS_ALL)	COV(PBS,PBD* MOUTHS_ALL)	VAR(PBS)	COV(PBD,PBD* MOUTHS_ALL
	COV(PBS, PBB)	COV(PBD, AGE1)	COV(PBD,PBD* MOUTHS_ALL)	VAR(PBD)
	COV(PBS, PBD)	COV(PBD* MOUTHS_ALL, AGE1)	VAR(PBD* MOUTHS_ALL)	COV(PBD* MOUTHS_ALL, PBB)
	COV(INCOME, PBD* MOUTHS_ALL)	COV(PBD* MOUTHS_ALL MOUTHS_ALL)	COV(PBS,PBD* MOUTHS_ALL)	VAR(PBB)

This model uses indicator variables for age and observed air lead.

Table 5-2. Comparison of Structural Equation Models
BLOOD LEAD EQUATION

Variables	GLS		AGLS	
	Kellogg	E. Helena	Kellogg	E. Helena
SOIL LEAD	0.000^{CON}	2.120***	0.341*	2.236***
	(0.039)	(0.565)	(0.160)	(0.786)
DUST LEAD	1.268*	0.793^{+}	1.220*	0.599^{NS}
	(0.588)	(0.433)	(0.582)	(0.404)
DUST LEAD* MOUTHS_ALL	0.847^{+}	0.814**	0.922**	0.940***
	(0.453)	(0.309)	(0.361)	(0.310)
AGE1	-5.328**	-0.489^{NS}	-5.717***	-0.337^{NS}
	(1.985)	(0.898)	(1.773)	(1.024)
AIR LEAD	36.16***	0.000^{CON}	21.90***	0.000^{CON}
	(11.00)	(0.018)	(7.08)	(0.021)
INCOME	-1.035^{NS}	0.776^{NS}	-1.046^{NS}	0.445^{NS}
	(0.773)	(0.753)	(0.826)	(0.592)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10
CON means that the estimate was constrained at its lower bound.

Asymptotic standard errors in parentheses.

This model uses indicator variables for age, and uses observed air lead.

Table 5-3. Comparison of Structural Equation Models
SOIL LEAD EQUATION

Variables	GLS		AGLS	
	Kellogg	E. Helena	Kellogg	E. Helena
AIR LEAD	28.45***	0.290***	27.24***	0.280***
	(5.87)	(0.026)	(5.23)	(0.033)
DUST LEAD EQUATION				
SOIL LEAD	0.097*	0.892***	0.091^{+}	0.884***
	(0.041)	(0.101)	(0.054)	(0.135)
AREA1	-7.758**	-0.274^{NS}	-0.066^{NS}	-0.544^{NS}
	(2.96)	(0.614)	(0.498)	(0.552)
AREA2	52.98***	0.290^{NS}	8.05***	0.325^{+}
	(16.80)	(0.212)	(2.78)	(0.186)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10

Asymptotic standard errors in parentheses. CON means that parameter estimate is constrained to be non-negative.

This model uses indicator variables for age, and uses observed air lead.

Table 5-4. Comparison of Structural Equation Models
DUST LEAD * MOUTH_ALL

Variables	GLS		AGLS	
	Kellogg	E. Helena	Kellogg	E. Helena
DUST LEAD	0.908***	1.088***	0.861***	1.925***
	(0.034)	(0.043)	(0.047)	(0.057)
MOUTHS_ALL	1.987***	1.256***	1.969***	1.086***
	(0.094)	(0.099)	(0.139)	(0.129)
AGE1	-0.149^{NS}	-0.246^{NS}	-0.143^{NS}	-0.196^{NS}
	(0.184)	(0.161)	(0.115)	(0.205)
INCOME	0.024^{NS}	0.124^{NS}	-0.014^{NS}	0.100**
	(0.071)	(0.130)	(0.047)	(0.037)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10

Asymptotic standard errors in parentheses.

This model uses indicator variables for age, and uses observed air lead.

Table 5-5. Comparison of Structural Equation Models
MEASUREMENT STANDARD DEVIATIONS

Variables	GLS		AGLS	
	Kellogg	E. Helena	Kellogg	E. Helena
BLOOD LEAD (μg/dL)	6.13 (0.126)	3.70 (0.102)	6.05 (0.137)	3.60 (0.129)
DUST LEAD (μg/g)	1315 (0.123)	748 (0.097)	1157 (0.298)	661 (0.155)
SOIL LEAD (μg/g)	3415 (0.128)	511 (0.097)	2190 (0.330)	537 (0.159)
DUST LEAD* MOUTHS_ALL (μg/g)	583 (0.122)	641 (0.099)	590 (0.199)	540 (0.249)

Relative standard error of variance shown in parentheses.

This model uses dummy variables for age, and uses observed air lead.

had been a sufficiently dense air sampling network to allow estimation of air leads specific to each house. It is obvious that air lead in the Idaho study is a surrogate variable for other site-specific differences. Further exploration of these data have not identified any variables that do a better job of explaining these differences. Since the AGLS models presented in Table 5-2 produce a reasonably consistent set of hand dust lead estimates and very non-significant P values, we will recommend use of this model specification.

The soil and dust lead pathway equations in Table 5-3 are also the most plausible of those models developed. The soil lead vs. air lead coefficient for East Helena is plausible; that for Kellogg is not, but the model is retained for purposes of comparison. The dust lead models are particularly interesting. In this case, the AGLS estimates for dust lead vs. soil lead and dust lead vs. air lead are positive and statistically significant (one-tailed). The dust lead vs. soil lead coefficient is much lower around the inactive smelter in Kellogg, but the air lead coefficient is much higher. In view of the problems in interpreting the soil and blood lead vs. air lead coefficients, we cannot comment further on the dust lead coefficient either. It is possible, however, that the difference in dust lead vs. soil lead coefficients is indicative of a real difference in the relative importance of environmental lead pathways in these two communities. The inclusion of a dummy variable for AREA1 does not much affect this conclusion.

In Table 5-4, we show that the cross-product of dust lead and mouthing frequency has a strong linear correlation with its linear components, dust lead and mouthing. Neither age nor income significantly improve the model for this surrogate hand lead variable, which is not perfectly collinear with dust lead. This constructed variable should therefore have many of the statistical properties of real-life hand lead data, and should be a useful surrogate index of proximate exposure to dust lead.

Estimates of the surrogate variable and measurement uncertainty are shown in Table 5-5. Note that estimates of the blood lead standard deviation (6.0 μg/dl in Kellogg and 3.6 μg/dl in East Helena) are much larger than the stated analytical errors, thus reflect differences among individual children in lead exposure and in biokinetics of lead absorption and elimination. The standard deviations in soil lead and dust lead concentrations reflect differences in lead pathways that are not adequately characterized by the composited exterior soil and interior dust measurements. These are large for the Kellogg data (2190 μg/g for soil lead, 1157 μg/g for dust lead). The environmental lead data for East Helena have much less idiosyncratic variability from one house to another, only 537 μg/g for soil lead and 661 μg/g for dust lead. This shows that around the active lead smelter in East Helena, soil and dust lead levels remain correlated from the major point source of exposure. The greater surrogate measurement variability in Kellogg suggests that during the two years in which the Bunker Hill smelter was inactive, other human activities continued to modify the household soil and dust lead levels. The uncertainty associated with our hand lead surrogate is much smaller than that of dust lead, 590 μg/g in Kellogg and 540 μg/g in East Helena. The constructed hand lead variable appears to be a very useful index of exposure.

6.0 COMPARISON WITH THE MIDVALE STUDY

6.1 MATERIALS AND METHODS

The 1983 CDC studies were the most directly comparable of the studies whose data were available to us. We were also interested in the extent to which these older studies could be compared with a recent study using somewhat different variables. The Midvale study was carried out in a Utah community which had formerly been the site of extensive metal smelting and processing operations (Bornschein et al. 1990). Elevated levels of lead were still found in soil and dust near the industrial sites. Data were provided to U.S. EPA by Dr. Bornschein and his associates, and EPA provided the data to us. Extensive preliminary analyses were performed on this data set (Marcus and Bernholc 1990). We report here only those results that are most directly comparable to the methodological studies in Sections 4 and 5. The more detailed models included an additional equation for paint lead loadings predicted by location, housing age, and the recent occurrence of paint removal (interior or exterior). The simpler models here are based on the following structure:

location + house age + paint removal --> soil lead

soil lead + paint removal + SES --> dust lead

dust lead + mouthing behavior + age + SES --> hand lead
(surrogate measure is dust lead X mouthing frequency)

dust lead + hand lead + soil lead + age + SES + paint removal --> blood lead.

The sample size here was relatively small, only N = 135 for whom essentially complete environmental and biological data were available. The mouthing behavior variable was also different, more directly representing the frequency of mouthing behavior than those items used in the CDC interviews. The soil lead variable we used for these analyses was not a composite of front

and back yard soil samples, but a composite of the soil samples taken at the dwelling unit yard perimeter.

6.2 Results

The goodness of fit summary in Table 6-1 shows that the AGLS model provides a satisfactory fit (P = 0.12) to the data. The extra degrees of freedom are related to the much more detailed information on location of individual households, and on the additional information on effects of paint removal.

The AGLS blood lead regression model in Table 6-2 is somewhat different than those derived for the Kellogg and East Helena children. The first difference is that the soil lead regression coefficient is large (3.05 μg/dl per mg/g lead in perimeter soil, whereas the interior dust lead coefficent is constrained at 0. The regression coefficient for dust lead X mouthing frequency is comparable to the sum of the two dust lead coefficients in the other models, about 1.55 μg/dl per mg/g lead in interior dust for a child with average mouthing behavior. The interview questions and frequency coding for mouthing behavior used by the Cincinnati investigators in Midvale appear to provide a more effective characterization of the hand-to-mouth pathway for dust lead than does the analogous index we used for the Kellogg and East Helena children. The AGLS model also shows a marginally significant increase in blood lead from paint lead removal (about 0.7 μg/dl even long after the abatement). Children at age 6 to 17 months (AGE1 = 1) had significantly lower blood leads than older children, and those with higher SES also had significantly lower blood leads.

The soil lead model in Table 6-3 shows a significant quadratic response surface with significant "ridge" of elevated soil lead towards the northeast suggesting a relatively localized source. Post-war housing had significantly lower perimeter soil lead levels, about 292 μg/g. Nineteenth century housing had slightly (but non-significantly) increased soil leads. The dust lead model in Table 6-4 shows that soil lead is probably a major

Table 6-1. Comparison of Structural Equation Models
GOODNESS OF FIT

	Midvale	
	GLS	AGLS
Loss Function	0.29881	0.18856
Chi-Squared	80.379	50.722
DF	40	40
P	<0.001	0.11922
Largest Residuals in Variance or Covariance	VAR(PBB)	VAR(PBD)
	COV(PBB, PBD)	VAR(PBS)
	VAR(X)	VAR(AGE1, RMVPNT)
	VAR(X, PBD*MOUTHS_ALL)	VAR(PBB)
	COV(PBB, XSQR)	COV(SES, MOUTHS_ALL)
	VAR(PBB)	COV(PBS, AGE_SQ)

Table 6-2. Comparison of Structural Equation Models
BLOOD LEAD EQUATION

Variables	Midvale	
	GLS	AGLS
SOIL LEAD	2.199***	3.047***
	(0.711)	(0.889)
DUST LEAD	0.000[CON]	0.000[CON]
	(0.007)	(0.006)
DUST LEAD* MOUTHS_ALL	1.665***	1.553**
	(0.569)	(0.577)
AGE1	-1.378*	-1.744***
	(0.631)	(0.574)
SES	-0.0813***	-1.0587***
	(0.0246)	(0.0114)
REMOVE PAINT	0.651[NS]	0.676+
	(0.562)	(0.402)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10

Table 6-3. Comparison of Structural Equation Models
SOIL LEAD EQUATION

Variables	Midvale	
	GLS	AGLS
PRE 20	0.148+	0.101^{NS}
	(0.090)	(0.072)
POSTWW2	-0.339***	-0.292***
	(0.066)	(0.055)
X	-1.228***	-1.179***
	(0.304)	(0.295)
Y	0.468*	0.320**
	(0.202)	(0.126)
XSQR	2.158***	2.126***
	(0.770)	(0.602)
YSQR	-0.164^{NS}	-0.050^{NS}
	(0.124)	(0.067)
XY	-0.467^{NS}	-0.523^{NS}
	(0.438)	(0.379)
RMVPNT	-0.045^{NS}	-0.0038^{NS}
	(0.045)	(0.025)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10

Table 6-4. Comparison of Structural Equation Models
DUST LEAD EQUATION

Variables	Midvale GLS	Midvale AGLS
SOIL LEAD	0.717***	0.756***
	(0.091)	(0.077)
PRE20	0.060^{NS}	-0.075^{NS}
	(0.119)	(0.060)
POSTWW2	0.084^{NS}	0.060^{NS}
	(0.304)	(0.295)
REMOVE PAINT	0.075^{NS}	0.320**
	(0.060)	(0.027)
SES	-0.0094^{NS}	0.0109^{NS}
	(0.0280)	(0.0093)
DUST LEAD*MOUTHS_ALL		
DUST LEAD	0.539***	0.834***
	(0.056)	(0.060)
MOUTHS_ALL	0.548***	0.504***
	(0.034)	(0.043)
AGE1	0.013^{NS}	-0.004^{NS}
	(0.055)	(0.033)
SES	-0.011^{NS}	-0.0700*
	(0.021)	(0.030)

NOTE: Statistical significance (two-tailed)

*** means P < 0.001
** means 0.001 < P < 0.01
* means 0.01 < P < 0.05
+ means 0.05 < P < 0.10 (0.05 one-tailed)
NS means P > 0.10

source of dust lead, contributing about 76% of the household dust lead concentration. Paint removal has also significantly increased soil lead, with an after-effect averaging 83 μg/g. Neither SES nor housing age added significantly to household dust leads, apart from their correlation with these other factors. The hand lead surrogate (dust lead X mouthing) showed a strong linear dependence on dust lead and mouthing behavior, as we expected from the construction of this variable, and a modest dependence on SES.

The surrogate measurement errors estimated in Table 6-5 are much smaller than those in Sections 4 and 5. The more recent study appears to give much more relevant information about environmental lead pathways in children than do the earlier studies. We would optimistically attribute this to a combination of: (1) improved analytical methods for environmental and biological lead samples, (2) much more refined sampling protocols for soil, dust, and blood, and (3) improvements in survey questions for household risk factors and individual child behaviors that affect lead exposure.

Table 6-5. Comparison of Structural Equation Models
MEASUREMENT STANDARD DEVIATION

Variables	Midvale	
	GLS	AGLS
BLOOD LEAD	2.76 (0.100)	2.41 (0.138)
SOIL LEAD	238 (0.091)	155 (0.282)
DUST LEAD	306 (0.097)	205 (0.118)
DUST LEAD* MOUTHS_ALL	233 (0.100)	176 (0.195)

Relative standard error of variance shown in parentheses.

7.0 CONCLUSIONS AND DISCUSSION

The comparable structural equation regression estimates for three sites are shown in Table 7-1. We have chosen three rather different situations, even though all three are related to the processing of lead ores and the emission of leaded particulates into the environment of nearby communities. The Smelterville community near Kellogg had been the site of a primary lead smelter for nearly a century, including a period of time after a Sept. 1973 baghouse fire in which extremely high lead levels were found in the air, in soils and dusts, and in children. However, the Idaho smelter had been inactive for two years preceding the 1983 survey. The community had been made aware of the dangers of childhood lead poisoning for over a decade, and it is likely that substantial effort was still being put into household cleaning and dust control, and other measures intended to reduce contact of children with leaded dusts and soils. The East Helena smelter was still active in 1983, and had been operating at roughly the same level of activity for at least five years preceding the 1983 study. In Midvale, the lead smelting had ceased long ago, but there was still some concern about high levels of lead that remained in many household soils and in large piles of mine wastes and mill tailings. Many residents of Midvale were newcomers whose awareness of childhood lead hazards may not yet be fully developed.

The summary of results in Table 7-1 show a surprising amount of consistency, in spite of our efforts to develop models that were specific to risk factors in each community. Note that:

(1) The coefficients between blood lead and total dust lead are not significantly different among the three communities. The coefficients between blood lead and the surrogate hand lead variable (dust lead times frequency of mouthing behavior) are not significantly different among the three communities. This suggests that lead in fine household dust particles has roughly the same bioavailability, whatever the source of the household

dust lead (e.g. soil tracked into the house, particles deposited from the air, or fine fragments of peeling lead paint).

TABLE 7-1. STRUCTURAL EQUATION MODEL COMPARISONS FOR THREE STUDIES

Variables	BLOOD LEAD Kellogg	E. Helena	Midvale
SOIL LEAD	0.34 (0.16)	2.24 (0.79)	3.05 (0.89)
DUST LEAD +DUST LEAD* MOUTHS_ALL	2.14 (0.85)	1.54 (0.51)	1.55 (0.58)
DUST LEAD* MOUTHS_ALL	0.92 (0.36)	0.94 (0.31)	1.55 (0.58)
AGE1	-5.72 (1.77)	-0.34 (1.02)	-1.74 (0.57)
	DUST LEAD*MOUTHS_ALL (surrogate for hand lead)		
DUST LEAD	0.861 (0.047)	0.925 (0.057)	0.834 (0.060)
MOUTHS_ALL	1.969 (0.139)	1.086 (0.129)	0.504 (0.043)
	DUST LEAD		
SOIL LEAD	0.091 (0.054)	0.884 (0.135)	0.756 (0.077)

Estimated standard errors in parentheses.

(2) The direct coefficient from soil lead to blood lead is not significantly different for East Helena and Midvale children; it is significantly smaller for Kellogg children. This suggests that, under some circumstances, soil lead may be relatively less bioavailable at some sites than at others. We cannot identify any factors within the data set that may account for this difference. The fact that there is an indirect pathway from soil lead to dust lead to blood lead already been included in these models, so that this estimates a separate and distinct route of exposure to lead in soil other than the house dust pathway.

(3) Both direct and indirect pathways for exposure to soil lead are statistically significant at all sites.

(4) Blood leads are significantly lower for one-year olds (age 6 to 17 months) at Kellogg and Midvale, and lower at East Helena. The 1-year-old blood leads in the Idaho sample average 5.7 μg/dl lower than those of older children, even when adjusted for income and location. We may attribute this to the fact that the Bunker Hill smelter closed in November, 1981, and the sample was taken about 21 months later in August, 1983. The older children would have accumulated much higher body burdens of lead during the earlier period, and the excessive lead stored in the body would contribute to elevated blood leads in older children not seen in the 1-year-olds.

(5) The pathway from soil lead to dust lead is also significantly less at Kellogg than at E. Helena or Midvale. At Midvale and E. Helena, house dust lead concentration is 76 to 88 percent of the soil lead concentration near the house. This suggests that soil lead is a primary source of present dust lead levels in these communities. The ratio is much lower in Kellogg, only 9 percent. Adjustment for area differences may account in part for the lower ratio, but the result persists across a wide variety of model specifications and is likely to be real. One possible explanation is that the greater attention to house cleaning and more time for cleaning to take effect without

rapid house dust recontamination may largely "uncouple" dust lead and soil lead in Kellogg. Note that the blood lead vs. total dust lead coefficient is somewhat higher in Kellogg (2.14 μg/dl per mg/g dust lead, vs. 1.54-1.55 μg/dl per mg/g in the other communities), which partially offsets the lower soil lead coefficient (0.34 μg/dl per mg/g), suggesting some possibility of confounding between soil and dust hazards.

(6) The regression coefficients for (dust lead X Mouths_all) vs. dust lead are not significantly different, 0.834 to 0.925. This is not surprising, since the average mouthing frequency has been standardized to have mean value 1. The downward bias may reflect covariation between dust lead levels, mouthing behavior, age and socioeconomic status. Likewise, the significantly different linear regression coefficient for (dust lead X mouths_all) vs. (mouths_all) is roughly proportional to the mean dust leads in the three studies. We have used this response variable in place of hand lead loading, which was not available for these studies but will be available for most others. The non-linear variable we constructed is not perfectly collinear with dust lead, so that we may still use the structural equations approach. In fact, this variable is a relatively good proximate predictor of childhood blood lead -- better than dust lead alone.

The statistical method used to fit the models may strongly affect the statistical significance of the coefficients. This controls the structure of the model that is finally selected, if pathways are retained only when they have statistically significant coefficients. We do not recommend this practice. The physical and biological plausibility of the pathway should govern its inclusion or exclusion in the model, since the validity of nominal significance levels depends completely on using the models only to test pre-specified (causal) hypotheses. We have emphasized the most robust of the methods currently implemented in statistical programs for microcomputers, the AGLS (Arbitrary distribution Generalized

Least Squares) method that is known to be distribution-free in large samples. The AGLS models are in general consistent with models in earlier EPA documents, but provide a great deal of additional information on the relative importance of different environmental lead pathways.

We have shown that it is possible to develop very useful statistical models for comparison of potential childhood lead burdens from environmental lead in the home environment of the child. Similarity of data base variables and statistical analysis methods allow meaningful inclusion of site-specific differences in lead exposure. Fortunately, it may not be necessary to perform a site-specific survey of potentially affected populations in order to make a realistic assessment of risk. Our analyses showed that many parameters are similar from one site to another, in spite of differences in lead source and pattern of early childhood exposure. Additional studies in progress will allow us to identify those site-specific parameters that are critical for risk assessment. These are likely to include site-specific differences in soil-to-dust pathways, previous lead exposure history, and bioavailability of leaded dusts and soils. Site-specific factors that can be easily determined by rapid field data collection and laboratory analyses could be used to provide input parameters to generic predictive models, such as the EPA or CDC lead biokinetic models. This will facilitate the estimation of abatement effectiveness or soil cleanup levels using only the critical site-specific information. We have shown that some inter-site comparisons will require site-specific information on lead pathways.

REFERENCES

Bentler PM. 1989. EQS Structural Equations Program Manual. BMDP Statistical Software, Los Angeles, CA.

Bornschein RL, Succop P, Dietrich RN, Clark CS, Que Hee S, Hammond PB. 1985. The influence of social and environmental factors on dust lead, hand lead, and blood lead levels in young children. Environ Res 38:108-118.

Bornschein RL. 1989. Interaction of children with their environment: Age-related associations with lead exposure. Presented at: Lead: A Conference on Advances in Lead Research with Related Implications for Environmental Health. NIEHS, RTP, NC.

Bornschein RL, Clark CS, Dietrich KN, et al. 1989. Factors influencing longitudinal developmental blood lead profiles in children. Presented at "Lead: A Conference on Advances in Lead Research with Related Implications for Environmental Health" National Institute of Environmental Health Sciences, RTP, NC.

Bornschein RL, Clark CS, Grote J, Peace B, Roda S, Succop P. 1988. Soil lead - Blood lead relationshipin a former lead mining town. In: Environmental Geochemistry and Health, Monograph Series 4, Lead in Soil: Issues and Guidelines. (Eds) BE Davies and BG Wixson. Science Review Limited, Northwood, England. pp. 149-160.

Bornschein RL, Succop PA, Krafft RM, Clark CS, Peace B, Hammond PB. 1986. Exterior surface dust lead, interior house dust lead and childhood lead exposure in an urban environment. Reprinted from Trace Substances in Environmental Health, II. A symposium, DD Hemphill, Ed., University of Missouri, Columbia. MO.

Browne MW. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology 37:62-83.

CDC. 1983. East Helena, Montana Child Lead Study. Centers for Disease Control, Public Health Service, US Department of Public Health and Human Services, Atlanta, GA.

Chaney RL, Mielke HW, Sterrett SB. 1988. Speciation, Mobility and Bioavailability of Soil Lead. In: Environmental Geochemistry and Health, Monograph Series 4, Lead in Soil: Issues and Guidelines. (Eds) BE Davies and BG Wixson. Science Review Limited, Northwood, England. pp. 105-130.

Clark S, Bornschein RL, Succop P, Peace B, Ryan J, Kochanowski A. 1988. The Cincinnati soil-lead abatement demonstration project. In: Environmental Geochemistry and Health, Monograph Series 4, Lead in Soil: Issues and Guidelines. (Eds) BE Davies and BG Wixson. Science Review Limited, Northwood, England. pp. 287-300.

Clark S, Bornschein R, Ryan J, Dochanowski A, Succop P, Peace B. 1988. The Cincinnati Soil-Lead Abatement Demonstration Project. Presented at Lead in Soil Conference, Chapel Hill, NC, March 7-9, 1988. Will appear in special issue of Environ Geochem and Hlth.

Cohen J, Marcus A, Elias R. 1990. Estimating childhood multi-media lead exposure: Expanded exposure/uptake/biokinetic model. For presentation at the 83rd Annual Meeting & Exhibition, Air & Water Management Assoc. Pittsburgh, PA. June 24-29, 1990.

Colorado Department of Health (Division of Disease Control and Environmental Epidemiology), University of Colorado at Denver (Center for Environmental Sciences), and US Dept of Health and Human Service (ATSDR/PHS). Leadville Metals Exposure Study. April 1990.

Farrell K. 1988. Baltimore soil-lead abatement demonstration project. In: Environmental Geochemistry and Health, Monograph Series 4, Lead in Soil: Issues and Guidelines. (Eds) BE Davies and BG Wixson. Science Review Limited, Northwood, England. pp. 281-285.

Fuller WA. 1987. Measurement Error Models. John Wiley & Sons, Inc., Somerset, NJ.

Johnson T., Wijnberg L. 1988. Statistical analyses of lead exposure data collected in East Helena, Montana. PEI Associates, Durham, NC (in preparation).

Kupper LL. 1984. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. Am J Epidemiol 120:643-648.

Lewis and Clark County Health Department et al. 1986. East Helena, Montana: Child Lead Study. Summer 1983. July 1986.

Marcus AH, Holtzman AP. 1988b. Relationships between infant blood lead concentrations and lead in water or liquid diet. Report from Battelle Columbus Division to Office of Drinking Water, USEPA. Contract No. 68-02-4246. April 1988.

McIntyre D, Mahoney M. 1988. Boston soil lead project. In: Environmental Geochemistry and Health, Monograph Series 4, Lead in Soil: Issues and Guidelines. (Eds) BE Davies and BG Wixson. Science Review Limited, Northwood, England. pp. 247-252.

Panhandle District Health Department et al. 1986. Kellogg Revisited -- 1983: Childhood blood lead and environmental status report. July 1986.

Rosen JF, Markowitz ME, Bijur PE, Jenks ST, et al. 1990. Sequential measurements of bone lead content by L-X-Ray flourescence in $CaNa_2$ EDTA-treated lead-toxic children. Environ Hlth Perspec, November 1990 (in press).

Schilling RJ, Bain RP. 1988. Predicting children's blood lead levels on the basis of household-specific soil lead levels. Am J Epidemiology 3:197-205.

Steele MJ, Beck BD, Murphy BL, Strauss HS. 1990. Assessing the contribution from lead in mining wastes to blood lead. Reg Toxicol Pharm 11:158-190.

USEPA. Review of the National Ambient Air Quality Standards for Lead: Exposure Analysis Methodology and Validation. USEPA Office of Air Quality Planning and Standards, RTP, NC. Report No. EPA-450/2-89/011.

USEPA. Air Quality Criteria for Lead Volume I - IV. Environmental Criteria and Assessment Office, Office of Research and Development, Research Triangle Park, NC. EPA 600/8-83-028 a-d, June 1986.

ACKNOWLEDGEMENTS

We are most grateful to Dr. Edmond Maes Centers for Disease Control for providing the East Kellogg data sets, and for working with us on the analyses.

We are grateful to Drs. Robert Bornschei Clark of the University of Cincinnati for their valua and comments on the Cincinnati, Telluride and Midvale